SYSTEM AND METHOD FOR RESCORING N-BEST HYPOTHESES OF AN AUTOMATIC SPEECH RECOGNITION SYSTEM

ABSTRACT OF THE DISCLOSURE

A system and method for rescoring the N-best hypotheses from an automatic speech recognition system by comparing an original speech waveform to synthetic speech waveforms that are generated for each text sequence of the N-best hypotheses. A distance is calculated from the original speech waveform to each of the synthesized waveforms, and the text associated with the synthesized waveform that is determined to be closest to the original waveform is selected as the final hypothesis. The original waveform and each synthesized waveform are aligned to a corresponding text sequence on a phoneme level. The mean of the feature vectors which align to each phoneme is computed for the original waveform as well as for each of the synthesized hypotheses. The distance of a synthesized hypothesis to the original speech signal is then computed as the sum over all phonemes in the hypothesis of the Euclidean distance between the means of the feature vectors of the frames aligning to that phoneme for the original and the synthesized signals. The text of the hypothesis which is closest under the above metric to the original waveform is chosen as the final system output.